DOCUMENT RESUME

ED 175 811 SP 014 147

AUTHOR Reavis, Charles A.; Shine, Stephen C.
TITLE A Study of the Effects of Prefatory Remarks on Teacher Evaluation.
PUB DATE 77
NOTE 19p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Attitudes; *Bias; Evaluation Criteria; *Expectation; *Observation; *Predictive Validity; Teacher Behavior; *Teacher Evaluation

ABSTRACT

A test was made to determine the impact of positive or negative expectations on individuals conducting formal observation of teachers in the classroom for the purpose of evaluation. Two groups of graduate students who were receiving training in classroom observation viewed a video tape of a teacher instructing her class. Group one was given the impression that she had been a fine education student in college; group two was told casually that she had not been outstanding in college. A review of the evaluation ratings by the two groups revealed that their responses to the teacher's classroom behavior was colored by the prefatory remarks they had heard, and that they were inclined to judge her with some positive or negative bias. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED175811

#137

DEC 1977

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. A. Reavis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A STUDY OF THE EFFECTS OF PREFATORY REMARKS ON TEACHER EVALUATION

Category - Manuscript

Submitted by:

Dr. Charles A. Reavis
Professor of Educational
Administration & Supervision
College of Education
Texas Tech University
Lubbock, Texas 79409
(806) 742-2365

SP014147

Dr. Charles A. Reavis
Stephen C. Shine

A STUDY OF THE EFFECTS OF PREFATORY REMARKS ON TEACHER EVALUATION

The increasing costs of education, coupled with a drop in the Scholastic Aptitude Test scores have increased the public's demands that the schools should be held more accountable. These developments have received much treatment in the professional literature (Harnischfiger and Wiley, 1976: 117-119) and (Tyler, 1975: 38) and the press (Ross, 1977: 12-15). The back to basics movement can clearly be associated with this trend, as can mandated teacher evaluation from state legislatures (Bennett, 1974: 485-487). The public is demanding that school officials critically appraise not only their programs, but their teachers as well.

Just a few years ago, teacher evaluation was a casual affair. Recently, however, teachers have been fired for poor performance in the classroom, and there is acceptance within the profession itself that evaluation should have consequences, i.e., training programs to ameliorate noted teaching defects. In this atmosphere, the importance of the evaluation's rendering an accurate, unbiased assessment can hardly be overstated. If the teacher is fired and seeks legal counsel, the evaluation will be one of the standards used in the trial; if programs of remediation are to be implemented, the evaluator as a first requisite must have an accurate picture of the teacher's classroom behavior in order to plan an appropriate improvement program. However, if the evaluator has been cued to look for "problems" (even worse, if the supervisor has been

specifically cued to look for the teacher's problems in discipline, for example) it is likely that this is what he will attend to, even though there are other, more subtle factors that are critical to the teaching act, perhaps even underlying the one to which his attention has been directed.

It may be, on the other hand, that supervisors and other highly qualified staff do not allow themselves to be biased in such a way. They are professionals with years of experience in the observation of teachers. They have perhaps attended workshops and conferences on the observation and evaluation of teachers. Quite likely in their graduate work they have had the opportunity to sharpen their observation skills and study the influence of the expectancy effect on perception. Upon being cued to a certain problem a caution light may, in effect, flash and they may become particularly careful to observe all classroom behaviors.

The purpose of this study, therefore, was to determine if a prefatory remark did indeed influence the evaluation of a teacher's classroom performance.

A review of the literature indicated that a number of studies had been conducted about the expectancy effect. In brief the expectancy effect holds that more often than not subjects will see or obtain the results they have been led to anticipate. Studies involving concrete dependent variables, such as timing rats running a maze, and more abstract dependent variables such as an increase in I.Q. have confirmed the power of the expectancy effect.

Rosenthal and Fode (1963) found that psychology students who believed they had maze bright rats obtained significantly better times

than students who believed they had maze dull rats. Actually, all rats were equally skilled in the maze experiment. Moving to the classroom, Rosenthal and Jacobson (1966) led teachers to believe that certain pupils would make a learning spurt that year. Actually, the pupils had been selected from a table of random numbers. These pupils, however, did make learning and I.Q. gains, particularly those in the lower grades.

Elashoff and Snow (1971) reexamined the Rosenthal and Jacobson data of 1966. Elashoff and Snow felt that significant deficiencies existed, especially in the statistical treatment given the data by Rosenthal and Jacobson. Specifically, they were critical of the amount of randomization given the students, of the reporting of the level of significance, and of the non-reporting of rare score data.

Rosenthal and Rubin (1971) rebutted the Elashoff and Snow criticisms. They cited several statisticians who confirmed their original methodology.

In a study similar to Rosenthal-Jacobson, Cahen (1966) found that the type and amount of information given student teachers about pupils affected their scoring of pupils' papers in the predicted direction. Psychology students administering the vocabulary section of the Wechster Intelligence Scale for Children were told their subjects had high reading ability. Other psychology students were told their subjects had low reading ability. Actually, all subjects were of similar ability. The psychology students who believed they had brighter children scored them significantly higher on the test; those who believed they had low ability students scored them significantly lower. Rosenthal (1973) summarizing the results of 242 studies conducted on the induced expectancy effect reported 84 studies had significant results, about six times what would be expected by chance.

In summary, the studies conducted by Rosenthal, et al., (1963, 1966, 1971, 1973) have indicated that anticipation of results can have a definite

effect on the rating of students or experimental subjects. The effect appears to be related to the type of information given as Simon (1969) found when experimenters gave better scores to subjects who were identified as above average than to those who were supposedly below average. The study by Cahen (1966) also indicated that prior information altered the scores of experimental subjects in the predicted directions. No studies were found that treated the effect of biasing remarks on an evaluator's rating of a teacher's behavior in the classroom.

THE EXPERIMENTAL PROCEDURE

The study involved a total of thirty-four graduate students enrolled in an instructional assessment class during three different semesters. The students, both men and women, ranged fron those pursuing master's degrees to those pursuing a certificate of advanced study (thirty hours beyond the master's), either in educational administration or supervision. All of the students but one, a military officer, were either elementary or secondary classroom teachers, full time graduate students on leave from classroom duties, or administrative personnel from various public school systems. See Table 1.

Table 1

Population Included in Study

Degree in progress - N = 34	
Master's	20
Advanced certificate of study	14
Professional responsibilities - N = 34	
Teacher	21
Administrator, supervisor	12
Military	1

The students had just completed nine classroom hours of study explaining the influence of biases on a supervisor's rating of a teacher. The students were informed that the study of biasing effects was completed, and attention would now be turned to examining various instruments of evaluation.

The students were divided into groups I and II. Group I was to the left of the middle row; group II to the right. The students were told the reason for this was that the television screen was small and they would need to see it very clearly for the next lesson. One group remained in the classroom to view the video-tape; the other was directed to another classroom for a group discussion. Members of each group were instructed not to converse with one another as they viewed the video-tape or with members of the other group if they should meet during the change of rooms. The reason given for this was that it would violate confidentiality, an important supervisory characteristic. The actual reason for the no discussion rule was to guard against talk that would dilute the effects of the prefatory remark.

While group II met to discuss another topic, group I viewed a video-tape of a student teacher instructing a sixth grade elementary class on the difference between fact and opinion as seen in advertising.

The first group of students to view the video tape was given a 72 item check list (Table 1) of teacher behaviors. The group was instructed to mark a check (√) if a particular behavior was good, a minus (-) if the instruction was bad, and a zero (0) if the instruction was either not present or neither good nor bad. As a side remark, the professor indicated that the student teacher whom they were to view was one of the best student teachers that the school of education had produced, and

that was why she did not mind being video taped. The professor did not emphasize the point, nor were repeated references to the ability of the student teacher made other than the one casual remark. The students were reminded of the no discussion rule and were requested not to discuss having seen the video-tape with the student teacher, should they see her. The students then viewed the 22 minute tape, while the professor was with the other group, and marked their responses during the tape.

When the tape was over, the first and second groups exchanged rooms. The second group was then given the same instructions about marking the form as the first group. The only difference was that the professor said the student teacher they were about to view was not one of the best student teachers that the school of education had sent out into the field. As in the first group, the prefatory remark was casual and seemed to be, "Oh, by the way." That was the only reference made about the teacher's ability. The second group viewed and marked their answers, using the same codes that the first group used.

The answer sheets were then collected, unsigned, but marked as to which group they were in. The data were than analyzed using the cross-tabulation procedure found in Statistical Package For The Social Sciences. The crosstabulation program calculated the chi-squared statistic which determined the significance level of the difference in responses of the two groups for each of the 72 items on the questionnaire.

After the significant items in the rating form were identified, an independent panel of four university professors individually rated the twenty-four significant items indicating a high degree of ambiguity, moderate ambiguity, or low ambiguity using the following standards:

High inference. An item subject to significant interpretation

by the rater and great use of his judgment and values.

Example: "Teacher has pleasing personality."

Moderate inference. Requires the use of some rater judgment, but likely fairly clear to most experienced evaluators.

Example: "Teacher utilizes student interest in lesson."

Low inference. With a little training, almost any evaluator would rate a given teacher behavior in the same way. Minimal amount of rater judgment required.

Example: "Teacher provides every child a chance to participate in class."

RESULTS

The purpose of this study was to determine if a single prefatory remark would influence the evaluation of a teacher's classroom performance. The data revealed that on 24 of 72 items the responses of the two treatment groups were significantly different at the .05 level of confidence or less. On the basis of chance alone only four significant differences would have been expected; the actual results were six times greater than chance. All differences were in the predicted direction; that is, the group receiving the positive prefatory remark gave the teacher more positive ratings; the group receiving the mildly negative remark gave the teacher more neutral or negative ratings.

An independent panel of university professors rated the significant items on a scale of low, moderate, or high ambiguity. The results of this rating also appear in Table 2. As judged by this panel, six items were considered low in ambiguity whereas only two items were considered

Table 2

Questions Rejecting the Null Hypothesis with Statistical Data

No.	Question	Ambiguity Rating	Raw Score-Teacher Rating: Prefatory Remark: Positive: Good (√)	Positive: Neutral/Not Present (0)	Positive: Bad (-)	Slightly Negative: Good (√)	Slightly Negative: Neutral/Not Present (0)	Slightly Negative: Bad (-)	Raw Chi Square	Degrees of Freedom	Level of Significance
1	Management of light, ventilation, heat, etc.	L-M	7	10	0	1	16	0	4.08	1	.04
6	Poise	M-H	17	0	0	10	4	3	8.81	2	.01
7	Voice	L-M	16	1	0	6	5	6	13.21	2	.001
8	Use of English	L	17	0	0	9	1	7	10.46	2	.005
9	Enthusiasm	M-H	17	0	0	8	4	5	12.24	2	.002
10	Provides good emotional environment	M	16	1	0	10	5	2	6.05	2	.04
11	Provides opportunities for controlled freedom (self-discipline	M	13	4	0	6	8	3	6.91	2	.03
17	Use of time	M	12	5	0	9	3	5	5.93	2	.05
18	Organization and care of supplies and equipment	L	12	5	0	7	6	4	5.41	2	.05
28	Is clear and understood	M	16	1	0	10	4	3	6.18	2	.04

30	Provides for learning at higher cognitive levels	M-H	13	4	0	5	9	3	8.48	2	.01
31	Includes learning experiences in the affective domain	M	12	5	0	6	8	3	5.69	2	.05
33	Encourages learners to ask questions	L	15	2	0	8	6	3	7.13	2	.02
40	Makes logical connection with previous knowledge	M	15	2	0	7	8	2	8.51	2	.01
42	Stresses logical reasoning processes	M	16	1	0	8	5	4	9.33	2	.01
43	Sustains lesson interest	L-M	17	0	0	11	3	3	7.28	2	.02
46	Evokes discussion and total learner participation	L	17	0	0	11	5	1	7.28	2	.02
47	Commends success and effort	L	16	0	1	10	4	3	6.38	2	.04
48	Corrects faults by commending virtues	L	8	9	0	2	11	4	7.80	2	.02
52	Attention to psychological principles of learning	H	14	2	1	7	5	5	6.28	2	.04
54	Encourages self-evaluation by learner	M	11	6	0	5	7	5	7.32	2	.02

57	Provides for continuing evaluative feedback during class period	M	16	1	0	7	9	1	10.92	2	.004
60	Starts effort-evoking interest	M	17	0	0	8	7	2	12.24	2	.002
69	Develops learner self-control and self-direction	H	12	5	0	4	10	3	8.66	2	.01

H = High ambiguity
M = Moderate ambiguity
L = Low ambiguity

to be high in ambiguity. The mode rating was moderate ambiguity, which was accorded to ten of the twenty-four significant items.

DISCUSSION

The present study was designed to test whether the expectancy effect that has been noted in other experiments would be operative in the evaluation of teaching. Subjects in this study were all professionals, over one-third with experience in the evaluation of teaching, and all having received nine class hours of instruction in possible biases in the rating of teacher effectiveness. With this population there was a substantial question as to whether biasing effects would indeed occur.

Each of three classes of educational professionals, a total of 34 subjects, was divided in two parts, shown a video-tape of a teacher teaching, and requested to complete a rating form. One of each of the groups received one positive comment on the teacher's effectiveness. The other of each of the groups in each class received one moderately negative comment. The chi-squared statistic was calculated for each of the 72 items on the rating form. Twenty-four of the 72 items were rated significantly different by the two treatment groups, six times what would be expected by chance. It was concluded that the expectancy effect does occur in the rating of teaching, even after nine class hours of instruction on biases in the rating of teaching.

The finding that an independent panel judged six of the items as low in ambiguity, whereas only two were judged high in ambiguity adds weight to the conclusion that the results obtained were a function of the prefatory remark, rather than the result of ambiguous items. Generally speaking, the more ambiguous the item, the less it is based on observed data and the more potential for the prefatory remark to have impact.

However, the finding that the prefatory remark had an impact, even on items rated low in ambiguity suggests just how influential the prefatory remark can be, even on trained professionals.

In a sense this study was a test of the relative influence of the halo effect and the expectancy effect. The halo effect is the tendency to attribute other positive/negative qualities to an individual having one positive/negative quality - physical attrativeness, for example (Dion, Berscheid, and Walster, 1972). The teacher on video-tape was physically attractive as rated by all class members (N = 34); yet the prefatory remark by the professor was sufficient to wash out the halo effect. That is, subjects in the mildly negative conditions rated her significantly lower, even though they had a favorable reaction to her physical attractiveness.

This study could be criticized in that the prefatory remark was uttered by a status leader, the professor, and the tendency of groups to accept comments by the status leader is well known (Torrence, 1955; Riecken, 1958). However, this condition was similar to actual practice in which the principal may well cue supervisors or peer observers to the problems of a particular teacher.

The length of the questionnaire doubtless contributed to the magnitude of the significance obtained. It is difficult to rate 72 items and accurately reflect each of them on a 22 minute lesson. In such a situation there is likely a tendency to be more influenced by a positive or moderately negative remark about the effectiveness of the teacher. Yet the form was recently used by a large metropolitan school district for five years. It is, therefore, an instrument-in-use, if not an entirely satisfactory instrument for research purposes. It should also

be noted that a greater number than chance of even the low ambiguity items such as "Encourages learners to ask questions," were also influenced by the prefatory remark, suggesting that the remark, not the nature and number of the items, was the controlling factor in the results.

Assuming that the findings can be generalized to field supervisors, the implication of this study was that judgments of a teacher's teaching ability can be altered by a prefatory remark. The prefatory remark in this research was very moderate. If more explicit value judgments from a third party were to be rendered to a supervisor, then an even greater bias might be imparted to the evaluator, resulting in an even greater alteration of a teacher's evaluation. If the schools in general, and the teacher in particular, are to be held accountable, then their evaluation should reflect their performance. The evaluator should, therefore, be completely aware of how others might influence his ability to perform a major duty, a bias free evaluation of the teachers for whom he is responsible.

Bennett, R. 1974. "Accountability Under the Stull Bill," Educational Leadership, 31: 485-487.

Cahen, L. 1966. "An Experimental Manipulation of the Halo Effect," Doctoral dissertation, Stanford University.

Dion, K., Berscheid, E., and Walster, E. 1972. "What is Beautiful is Good," Journal of Personality and Social Psychology, 24: 289-90.

Elashoff, J. and R. Snow. 1971. Pygmalion Reconsidered, Worthing, Ohio: Charles A. Jones Publishing Co.

Haberman, M. 1970. "The Relationship of Bogus Expectation to Success in Student Teaching (or Pygmalion's Illegitimate Son)," Journal of Teacher Education, 21: 69-72.

Nie, N. and others. 1975. Statistical Package For The Social Sciences, 2nd. ed. New York: McGraw-Hill, 218-248.

Rafferty, M. 1974. "American Education 1975-2000," The Future of Education: 1975-2000, ed. T. L. Hipple, Pacific Palisades, California: Goodyear Publishing Co., Inc., 159-180.

Riecken, H. W. 1958. "The Effect of Talkativeness on Ability to Influence Group Solutions of Problems," Sociometry, 21: 309-321.

Rothbart, M., S. Dalfen, and R. Barrett. 1971. "The Effects of Teacher Interaction," Journal of Educational Psychology, 66: 49-54.

Rosenthal, R. 1973. "The Pygmalion Effect Lives," Psychology Today, 7: 56-59.

Rosenthal, R. and D. Rubin. 1971. Pygmalion Reaffirmed, U. S., Educational Resources Information Center, ERIC Document Ed. 059 247.

Rosenthal, R. and K. Fode. 1963. "The Effects of Experimenter Bias on the Performance of the Albino Rat," Behavioral Science, 8: 183-189.

Rosenthal, R. and L. Jacobson. 1966. Teacher Expectation. Determinants of Pupil's IQ Gains," Psychological Reports, 19: 15-18.

Ross, S. 1977. "A Diploma That Means Something," Parade, June 12, 1977.

Simon, W. 1969. "Expectancy Effects in the Scoring of Vocal Items: A Study of Scorer Bias," Journal of Educational Measurements, 6: 183-189.

Torrence, E. P. 1955. "Some Consequences of Power Differences on Decision-Making in Permanent and Temporary Three-Man Groups. In A. P. Hare, E. F. Borgalta, and R. F. Boles, eds., Small Groups: Studies in Social Interaction. New York: Knopf, 482-91.

Tyler, R. 1975. "Where Learning Happens," The National Elementary Principal, 54: 38-42.